UNCLASSIFIED

AD 412934

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA

UNCLASSIFIED

ORC 63-3 (RR)
25 MARCH 1963

PROBLEMS OF STATISTICAL INFERENCE FOR BIRTH AND DEATH QUEUEING MODELS

by
R. W. Wolff

OPERATIONS RESEARCH CENTER

INSTITUTE OF ENGINEERING RESEARCH

UNIVERSITY OF CALIFORNIA - BERKELEY

PROBLEMS OF STATISTICAL INFERENCE FOR BIRTH AND DEATH QUEUEING MODELS

by

R.W. Wolff
Operations Research Center
University of California, Berkeley

25 March 1963 ORC 63-3 (RR)

This research was partially supported by the National Science Foundation, Grant G 21034, and partially by the National Institute of Health, Grant GM-09606-02 with the University of California.

ABSTRACT

A large sample theory is presented for birth and death queueing processes which are ergodic and metrically transitive. The theory is applied to make inferences about how arrival and service rates vary with the number in the system. Likelihood ratio tests and maximum likelihood estimators are derived for simple models which describe this variation. Composite hypotheses such as that the arrival rate does not vary with the number in the system are considered. A numerical example illustrating these results is presented which includes the testing of both true and false composite hypotheses.

PROBLEMS OF STATISTICAL INFERENCE FOR BIRTH AND DEATH QUEUEING MODELS

I. Introduction

The orientation in queueing theory is to start with a model and determine the behavior of a system from it. In applications, however, one has the problem of deciding what model is appropriate from observing the behavior of a system.

For queueing processes, the problem has been untouched in the literature except for Clark [3], while for Markov processes, much has been done. A large sample theory based on maximum likelihood for Markov processes has been developed in a monograph by Billingsley [1], while his paper [2] contains a large bibliography. The results parallel the classical large sample theory as presented in Wilks [6]. Birth and death queueing processes are also Markov processes, and the theory in Billingsley directly applies. In this paper, the theory is applied to make inferences about how arrival and service rates vary with the number in the system.

The birth and death processes considered here are restricted so as to be ergodic and metrically transitive. A likelihood function is developed based on observing a system continuously over some interval of length T. Relations which are needed to work out results for particular models are developed, and limit theorems from Billingsley are presented. Estimators and hypothesis tests are derived for three simple models, as well as their asymptotic properties. All the models considered will have only a finite number of parameters.

Of particular interest are composite hypotheses such as that the service rate does not vary with the number in the queue. A composite hypothesis can be tested by first nesting it in a more general hypothesis. Methods for doing this are discussed at several points in the paper. A numerical example

illustrating these results is then presented, which includes the testing of both true and false composite hypotheses.

More general models for the variation of arrival and service rates with the number in the system are considered in an Appendix, as well as other ideas for further research.

II. Definition and Characteristics of the Process

A queueing process with Poisson arrivals and exponential service is a birth and death Process. If the number of customers waiting in the queue plus the number in service is j, we say the system or process is in state E_j. When the system is in E_j, arrival and service rates will be denoted by λ_j and μ_j respectively. Hence, λ and μ may be functions of E_j, but given E_j, they do not vary with time. (They are time homogeneous.) If a customer arrives while the system is in state E_j, we have transition $E_j \rightarrow E_{j+1}$. Similarly, if a service is completed while in state E_j, we have transition $E_j \rightarrow E_{j-1}$. In all models, the variation of λ_j and μ_j with E_j will be expressed in terms of a finite number of parameters, $\underline{\theta}' = (\theta_1, \theta_2, \ldots, \theta_k)$.* We will assume the following throughout this paper:

1. $\mu_0 = 0$
2. $\mu_j > 0$ for $j = 1,2,\ldots$
3. $\lambda_j > 0$ for all integral $j < M$, where M may be infinite
4. $\lambda_j = 0$ for $j = M, M+1, \ldots,$ if M is finite
5. The mean recurrence time for each state is finite.

Assumptions 1-4 prevent the existence of two disjoint non-null irreducible sets of state That is, the process must be metrically transitive. Should metric transitivity

---

* In this paper, $\underline{\theta}$ denotes a column vector, while its transpose $\underline{\theta}'$, denotes a row vector.

not hold, the theory will apply to each irreducible subset of states considered separately. With the addition of Assumption 5, the process will be ergodic. Conditions on the $\{\lambda_j\}$ and $\{\mu_j\}$ for Assumption 5 to hold are known, but of little importance here since the dimensionality of the parameter space is not affected, and the mean recurrence times will obviously be finite for most applications. In order not to violate Assumption 5, we must choose models with the property that for some $\underline{\theta}'$ there exists a neighborhood such that any $\underline{\theta}'$ belonging to that neighborhood defines a process which is ergodic and metrically transitive. A queueing model which is not ergodic is the following:

$$\lambda_j = a(j)^2 \qquad j = 0,1,\ldots \qquad a,b > 0$$

$$\mu_j = bj \qquad j = 1,2,\ldots$$

where the unknown parameters are $\underline{\theta}' = (a,b)$.

Ergodicity and metric transitivity quarantee that large sample theory is meaningful. Suppose the system is in state E_j at $t = 0$. We represent the probability that the system is in state E_k at t by $P[E_k(t) \mid E_j(0)] = P_{jk}(t)$. Under the above assumptions, $\lim_{t\to\infty} P_{jk}(t) = P_k > 0$, independent of j , such that $\{P_k\}$ is the unique stationary distribution for the process, $\sum_{k=0}^{\infty} P_k = 1$. Other properties of birth and death processes are discussed in Feller [5].

III. Observing the System

It will be assumed here and in what follows that the system is observed continuously over some time interval T . During T , the times at which customers arrive and services are completed are observed. This is equivalent to considering the state of the system as a step function (not monotone) over time. Suppose one starts to observe the system at $t_1 = 0$ and finds the system in state E_j where $j = x_1$. Let the first transition, $E_j \to E_k$,

(an arrival or completion of service) occur at t_2 , where $k = x_2$. Let the sequence of transitions after t_2 occur at $t_3,\ldots,t_{n+1}$, $t_{n+2},\ldots$ and let the number in the system instaneously after t_i be x_i . If $t_{n+1} \leq T$ and $t_{n+2} > T$, then one has the sequence $(x_1,t_1),(x_2,t_2),\ldots,(x_{n+1},t_{n+1})$, where n is the number of transitions observed during T .

IV. Constructing the Likelihood Function

In order to construct the likelihood function for the above sequence, we will first consider a single transition, $x_i \rightarrow x_{i+1}$. Note that because transitions are single step, $|x_{i+1} - x_i| = 1$. Let $x_i = j$ and let $\tau_i = t_{i+1} - t_i$, the time the system remains in state E_j , upon entering at t_i .

That portion of the likelihood function due to a single transition is made up of two components, the density function of time in state E_j evaluated at τ_i ; and given τ_i , the probability that the transition at t_{i+1} is the one observed (an arrival or completion of a service). Given x_i , the probability that $\tau_i > \tau$ is simply the probability that both the next arrival and next service completion occur after τ . Since these events are independent,

$$P[\tau_i > \tau] = (e^{-\lambda_i \tau})(e^{-\mu_i \tau}) = \exp -(\lambda_i + \mu_i)\tau \quad .$$

Hence, the density function for τ_i is:

$$f(\tau_i) = (\lambda_i + \mu_i) \exp -(\lambda_i + \mu_i)\tau_i \quad .$$

Note that we are using λ_i and μ_i to refer to the state of the system E_{x_i} in $(t_i, t_{i+1}]$, and not to any particular transition. Strictly, one should use λ_{x_i} and μ_{x_i} , but this notation is needlessly cumbersome.

It is not difficult to show that $P[x_{i+1} = x_i+1 \mid \tau_i] = \lambda_i/(\lambda_i + \mu_i)$, independent to τ_i . Equivalently, the probability that the next transition is due to a service is $\mu_i/(\lambda_i + \mu_i)$.

Hence, we obtain the portion of the likelihood function due to a single transition, $dF_i = dF_i(x_{i+1}, \tau_i \mid x_i, \underline{\theta})$, in the following form:

$$
\begin{aligned}
dF_i &= [\lambda_i/(\lambda_i + \mu_i)](\lambda_i + \mu_i)\exp -(\lambda_i+\mu_i)\tau_i \\
&= \lambda_i \exp -(\lambda_i+\mu_i)\tau_i \qquad \text{for } x_{i+1} = x_i + 1 \\
&= \mu_i \exp -(\lambda_i+\mu_i)\tau_i \qquad \text{for } x_{i+1} = x_i - 1 \quad . \qquad (1)
\end{aligned}
$$

The correct form depends on what was actually observed (an arrival or a service).

Since the future behavior of a Markov Process depends only on the state of the system when last observed, the portion of the likelihood function due to a single transition, conditioned on the state of the system, is independent of the past history of the system. Hence, the likelihood function of the entire sequency of n transitions is simply the product of terms such as (1). In addition, there is the probability that the system started in state E_{x_1} . If the system is initially observed at a random moment along the real line, $P[x_1] = P_{x_1}$, the stationary probability. However, for large sample theory, the assumption of random entry is not critical. Further, the portion of the likelihood function due to transition $n + 1$ occurring after T is $P[\tau_{n+1} > T - t_{n+1}] = \exp -(\lambda_{n+1}+\mu_{n+1})(T-t_{n+1})$. Hence, the likelihood function for the entire set of observations, $L_n(\underline{\theta})$, is the following:

$$
L_n(\underline{\theta}) = P[x_1] \exp -(\lambda_{n+1}+\mu_{n+1})(T-t_{n+1}) \prod_{i=1}^{n} dF_i \quad . \qquad (2)
$$

It is convenient to deal with the log of the likelihood function $\ln L_n(\underline{\theta})$.

As $T \to \infty$, $n \to \infty$ with probability one, and for large sample theory, the first term in $\ln L_n(\underline{\theta})$, $\ln P[x_1]$, can be ignored. When this is done, $\ln L_n(\underline{\theta})$ can be put in the following very useful form:

$$\ln L_n(\underline{\theta}) = \sum_{j=0}^{\infty} u_j \ln\lambda_j + \sum_{j=1}^{\infty} d_j \ln\mu_j - \sum_{j=0}^{\infty} r_j(\lambda_j + \mu_j) \quad , \tag{3}$$

where u_j is the number of transitions $E_j \to E_{j+1}$ (up) , d_j is the number of transitions $E_j \to E_{j-1}$ (down) , and r_j is the total time spent in E_j during T .

V. Useful Relations for Later Application

In what follows λ_j and μ_j will be allowed to vary with the number in the system in a way that can be functionally described in terms of a finite number of independent parameters. Let $\underline{\theta}' = (\theta_1,\ldots,\theta_k)$ be the vector representation of the parameters in Euclidean k-space. We have in general $\lambda_j = h_j(\underline{\theta})$, $\mu_j = g_j(\underline{\theta})$, suitably constrained as in Section II.

In addition to the stationary distribution, $\{P_j\}$, we will need the distribution $\{\pi_j\}$, where π_j is the probability that a randomly selected transition was out of state E_j . With greater precision, define $\pi_j = \lim_{T\to\infty} (u_j + d_j)/n$, which, for processes of the type considered here exists with probability one. For an ergodic process, a limit such as π_j is independent of the starting state, or equivalently, of any distribution of starting states. In evaluating this limit, however, it is conveneint to assume starting conditions are determined by the stationary distribution, $\{P_j\}$. When this is done, we obtain for any interval, T , $E[(u_j + d_j) \mid T] = (\lambda_j + u_j)P_jT$ and $E[n \mid T] = \sum_{j=0}^{\infty} (\lambda_j + \mu_j)P_jT$. It is now clear that

$$\pi_j = \frac{(\lambda_j + \mu_j)P_j}{\sum_{j=0}^{\infty} (\lambda_j + \mu_j)P_j} \quad .$$

The average rate of arrivals must equal the average rate of departures which we define as the average throughput, $R = \sum_{j=0}^{\infty} \lambda_j P_j = \sum_{j=1}^{\infty} \mu_j P_j$. Thus,

$$\pi_j = \frac{(\lambda_j + \mu_j)P_j}{2R} \qquad (j = 0,1,\ldots) \quad . \tag{4}$$

For birth and death models we also have the well-known relation:

$$\lambda_j P_j = \mu_{j+1} P_{j+1} \qquad (j = 0,1,\ldots) \quad . \tag{5}$$

For any specific model we can use these relations to determine the $\{P_j\}$ and the $\{\pi_j\}$ as needed.

Ignoring end effects, $\ln L_n(\underline{\theta}) = \sum_{i=1}^{n} \ln dF_i(\underline{\theta})$. From (1), dF_i has the following form:

$$\begin{aligned}(dF_i \mid x_i = j) &= \lambda_j(\underline{\theta}) \exp -(\lambda_j(\underline{\theta})+\mu_j(\underline{\theta}))\tau_i \qquad \text{if} \quad x_{i+1} = x_i + 1 \\ &= \mu_j(\underline{\theta}) \exp -(\lambda_j(\underline{\theta})+\mu_j(\underline{\theta}))\tau_i \qquad \text{if} \quad x_{i+1} = x_i - 1 \quad .\end{aligned}$$

Define

$$\begin{aligned}S_{nu} &= \frac{\partial}{\partial\theta_u} \ln L_n(\underline{\theta}) \qquad u = 1,2,\ldots,k \\ \underline{S}'_n &= (S_{n1},\ldots,S_{nk}) \\ G_u &= \frac{\partial}{\partial\theta_u} \ln dF_i \\ \underline{G}' &= (G_1,\ldots,G_k) \\ G_{uv} &= \frac{\partial^2}{\partial\theta_u \partial\theta_v} \ln dF_i \quad .\end{aligned}$$

Under suitable regularity conditions,* and from $\int_{\Omega} dF_i(x_{i+1}, \tau_i \mid x_i, \underline{\theta}) = E(1) = 1$ where Ω denotes the entire sample space, the following well-known results are obtained:

$$E(G_u) = 0 \tag{6}$$

and

$$E(G_{uv}) = -\sigma_{uv} \quad . \tag{7}$$

From (7), we see that the variance-covariance matrix of G_u, $E(\underline{GG}')$ is:

$$E(\underline{GG}') \equiv \sigma(\underline{\theta}) = (\sigma_{uv}) = (-E(G_{uv})) \quad . \tag{8}$$

VI. Limit Theorems from Billingsley [1]

THEOREM 1: [1] The asymptotic distribution of $n^{-1/2}\, \underline{S}_n$ as $T \rightarrow \infty$ is the multivariate normal with mean $\underline{\theta}$ and variance-covariance matrix $\sigma(\underline{\theta})$. We write:

$$n^{-1/2}\, \underline{S}_n \xrightarrow{L} N(\underline{\theta}\, , \sigma(\underline{\theta})) \quad .$$

Next consider the maximum likelihood estimator, $\hat{\underline{\theta}}'_n = (\hat{\theta}_{n1}, \ldots, \hat{\theta}_{nk})$, found by setting $\underline{S}_n = \underline{0}$. Assume that for large n the estimator is unique.

THEOREM 2: [1] The maximum likelihood estimator, $\hat{\underline{\theta}}$, converges in probability as $T \rightarrow \infty$ to the true vector of the parameters, $\underline{\theta}$, such that:

$$n^{1/2}(\hat{\underline{\theta}} - \underline{\theta}) \xrightarrow{L} N(\underline{0}\, , \sigma^{-1}(\underline{\theta})) \quad .$$

---

*Specifically, conditions on dF_i which permit the interchange of the order of the operations of differentiation and integration. See Cramer [4, p. 67 and 501], or Billingsley [1, p. 5 and 41].

If we hypothesize that $\underline{\theta}$ is some particular vector, $\underline{\theta}^o$ say, such that the probability distributions of all random variables in the process are completely specified, then we have a simple hypothesis. Given $H_0 : \underline{\theta} = \underline{\theta}^o$, the vector $\underline{\theta}$ is restricted to a single point, of dimensionality zero, in the k-dimensional parameter space. A general way to test such a hypothesis is with the likelihood ratio,

$$\frac{L_n(\underline{\theta}^o)}{L_n(\hat{\underline{\theta}}_n)}$$

or equivalently, with minus twice the log of this ratio:

$$2[\max_{\underline{\theta}} \ln L_n(\underline{\theta}) - \ln L_n(\underline{\theta}^o)] \quad . \tag{9}$$

If the value obtained from (9) exceeds some critical value, we reject H_0. We would like to know the asymptotic distribution of (9) when H_0 is true, and also under alternative hypotheses.

THEOREM 3: [1] If $H_0 : \underline{\theta} = \underline{\theta}^o$ is true, then minus twice the log of the likelihood ratio is asymptotically distributed as chi square with k degrees of freedom:

$$2[\max_{\underline{\theta}} \ln L_n(\underline{\theta}) - \ln L_n(\underline{\theta}^o)] \xrightarrow{L} \chi^2_k \quad .$$

Suppose that H_0 is false, and some alternative, $H_1 : \underline{\theta} = \underline{\theta}^1$, is true. Define the vector $\underline{h}$ as follows:

$$\underline{\theta}^o = \underline{\theta}^1 + n^{-1/2}\,\underline{h} \quad \text{or} \quad \underline{h} = n^{1/2}(\underline{\theta}^o - \underline{\theta}^1) \quad .$$

Now define the quadratic form, $\Delta^2 = \underline{h}'\sigma(\underline{\theta}^1)\underline{h}$, where Δ is the non-centrality

parameter in the non-central χ^2 distribution below.

THEOREM 4: [1] If $H_0 : \underline{\theta} = \underline{\theta}^o$ is false, but $H_1 : \underline{\theta} = \theta^1$ is true, then:

$$2[\max_{\underline{\theta}} \ln L_n(\underline{\theta}) - \ln L_n(\underline{\theta}^o)] \xrightarrow{L} \chi^2_{k,\Delta} \quad .$$

Suppose one imposes restrictions on the components of $\underline{\theta}$, such that some of them are equal, or are known, or are related to one another in such a way that given an appropriate proper subset of the $\{\theta_u\}$, the rest can be solved for uniquely. When such is the case, one can represent the system in terms of a vector of k' independent parameters, $\underline{\phi}$, such that $k' < k$. A model incorporating these new restrictions is said to be nested in the original model, and is represented by $\phi \subset \theta$. The components of $\underline{\phi}$ can be estimated and tested by methods used for $\underline{\theta}$ above. In addition, the hypothesis that the restricted model is true under the assumption that the original model is true can also be tested.

THEOREM 5: [1] If the restricted model is true, and the hypothesis that $\underline{\phi} = \underline{\phi}^o$ is also true, then:

$$2[\max_{\underline{\theta}} \ln L_n(\underline{\theta}^o) - \ln L_n(\underline{\theta}^o)] = 2[\max_{\underline{\theta}} \ln L_n(\underline{\theta}) - \max_{\underline{\phi}} \ln L_n(\underline{\phi})]$$

$$+ 2[\max_{\underline{\phi}} \ln L_n(\underline{\phi}) - \ln L_n(\underline{\phi}^o)] \quad ,$$

where $\underline{\theta}^o$ is determined by $\underline{\phi}^o$, and

$$2[\max_{\underline{\theta}} \ln L_n(\underline{\theta}) - \ln L_n(\underline{\theta}^o)] \xrightarrow{L} \chi^2_k$$

$$2[\max_{\underline{\phi}} \ln L_n(\underline{\phi}) - \ln L_n(\underline{\phi}^o)] \xrightarrow{L} \chi^2_{k'}$$

$$2[\max_{\underline{\theta}} \ln L_n(\underline{\theta}) - \max_{\underline{\phi}} \ln L_n(\underline{\phi})] \xrightarrow{L} \chi^2_{k-k'} \quad ,$$

where the last two expressions are asymptotically independent.

In the remainder of this paper, the theory is applied to birth and death problems of interest.

VII. Application of the Theory of Particular Models

MODEL 1: $\lambda_j > 0 \quad , \quad j = 0,1,\ldots,M-1 \qquad \lambda_M = 0$

$\mu_j > 0 \quad , \quad j = 1,2,\ldots,M \qquad \mu_0 = 0$

As mentioned earlier, restrictions must be placed on models we consider to explain the variation of the $\{\lambda_j\}$ and the $\{\mu_j\}$ with j so that there are only a finite number of unknown parameters. Model 1 accomplishes this by truncation, so that the number of the system cannot exceed M. There are $2M$ unknown parameters $\{\lambda_j\}$ and $\{\mu_j\}$ in all, which are required only to be finite and positive.

The maximum likelihood estimators are determined from Equation (3), which for this model has the form:

$$\ln L_n(\underline{\theta}) = \sum_{j=0}^{M-1} u_j \ln\lambda_j + \sum_{j=1}^{M} d_j \ln\mu_j - \sum_{j=0}^{M} \gamma_j(\lambda_j + \mu_j) \tag{10}$$

Differentiating with respect to each of the unknown parameters, we get:

$$\frac{\partial \ln L_n}{\partial \lambda_j} = \frac{u_j}{\lambda_j} - \gamma_j = 0$$

and therefore

$$\hat{\lambda}_j = \frac{u_j}{\gamma_j} \qquad \text{for } j = 0,1,\ldots,M-1 \tag{11}$$

$$\frac{\partial \ln L_n}{\partial \mu_j} = \frac{d_j}{\mu_j} - \gamma_j = 0$$

and therefore

$$\hat{\mu}_j = \frac{d_j}{\gamma_j} \qquad \text{for } j = 1,2,\ldots,M \quad . \tag{12}$$

There is nothing particularly surprising about (11) and (12). The estimate of λ_j is the number of arrivals while in state E_j divided by the total time spent in state E_j . The analysis does not stop here, however.

For the variance-covariance matrix of the estimators, we begin with Equation (1) and $G_u = \partial/\partial\theta_u \ln dF_i$, and given $x_i = j$ obtain:

$$G_{\lambda_j} = 1/\lambda_j - \tau_i \qquad G_{\mu_j} = \quad - \tau_i \qquad \text{for } x_{i+1} = j+1$$

$$G_{\lambda_j} = \quad - \tau_i \qquad G_{\mu_j} = 1/\mu_j - \tau_i \qquad \text{for } x_{i+1} = j-1$$

and all other $G_u = 0$.

The procedure to determine the variance-covariance matrix of the $\{G_u\}$, $\sigma(\underline{\theta})$, will be first to determine $\sigma(\underline{\theta} \mid x_i)$ for (7). Then, from $E(G_u \mid x_i) = E(G_u) = 0$ for all u , $E\{\sigma(\underline{\theta} \mid x_i)\} = \sigma(\underline{\theta})$. We continue to condition on $x_i = j$:

$$G_{\lambda_j\lambda_j} = -1/\lambda_j^2 \qquad G_{\mu_j\mu_j} = 0 \qquad \text{for } x_{i+1} = j+1$$

$$G_{\lambda_j\lambda_j} = 0 \qquad G_{\mu_j\mu_j} = -1/\mu_j^2 \qquad \text{for } x_{i+1} = j-1$$

and all other $G_{uv} = 0$.

As shown earlier, $P(x_{i+1} = j + 1 \mid x_i = j) = \lambda_j/(\lambda_j + \mu_j)$. Now

$$V(G_{\lambda_j} \mid x_i = j) = -E(G_{\lambda_j\lambda_j} \mid x_i = j) = (1/\lambda_j^2) \frac{\lambda_j}{\lambda_j + \mu_j} = \frac{1}{\lambda_j(\lambda_j + \mu_j)}$$

for $j = 0,\ldots,M-1$

$$V(G_{\mu_j} \mid x_i = j) = -E(G_{\mu_j\mu_j} \mid x_i = j) = (1/\mu_j^2) \frac{\mu_j}{\lambda_j + \mu_j} = \frac{1}{\mu_j(\lambda_j + u_j)}$$

for $j = 1,\ldots,M$, and $Cov(G_u,G_v \mid x_i = j) = -E(G_{uv} \mid x_i = j) = 0$ for all other (u,v) pairs. Also, $V(G_{\lambda_j} \mid x_i \neq j) = V(G_{\mu_j} \mid x_i \neq j) = 0$.

We now obtain:

$$V(G_{\lambda_j}) = E\{V(G_{\lambda_j} \mid x_i = k)\} = \sum_{k=0}^{M} V(G_{\lambda_j} \mid x_i = k)\pi_k = \frac{\pi_j}{\lambda_j(\lambda_j + \mu_j)}$$

for $j = 0,1,\ldots,M-1$

$$V(G_{\mu_j}) = E\{V(G_{\mu_j} \mid x_i = k)\} = \frac{\pi_j}{\mu_j(\lambda_j + \mu_j)}$$

for $j = 1,2,\ldots,M$, and $Cov(G_u,G_v) = 0$. Hence, using Equation (4),

$$V(G_{\lambda_j}) = \frac{P_j}{2\lambda_j R} \qquad \text{for } j = 0,1,\ldots,M-1$$

$$V(G_{\mu_j}) = \frac{P_j}{2\mu_j R} \qquad \text{for } j = 1,2,\ldots,M \quad .$$

Since all the covariances are zero, $\sigma(\underline{\theta})$ is a diagonal matrix. Let $\underline{G}' = (G_{\lambda_0},\ldots,G_{\lambda_{M-1}},G_{\mu_1},\ldots,G_{\mu_M})$, then $\sigma(\underline{\theta}) = (1/2R)(\ell_i\delta_{ij})$ where

$\ell' = (P_0/\lambda_0,\ldots,P_{M-1}/\lambda_{M-1},P_1/\mu_1,\ldots,P_M/\mu_M)$. Now $\sigma^{-1}(\underline{\theta}) = (\sigma^{uv}) = 2R(\delta_{ij}/\ell_i)$, and letting $\underline{\hat{\theta}}' = (\hat{\lambda}_0,\ldots,\hat{\lambda}_{M-1},\hat{\mu}_1,\ldots,\hat{\mu}_M)$, we have by Theorem 2:

$$n^{1/2}(\underline{\hat{\theta}} - \underline{\theta}) \xrightarrow{L} N(\underline{0}\ , \sigma^{-1}(\underline{\theta})) \quad .$$

Since $\sigma^{-1}(\underline{\theta})$ is a diagonal matrix, the maximum likelihood estimators are asymptotically independent. If we let $V_{as}(X)$ be the asymptotic variance (the variance of the asymptotic distribution) of X , then:

$$V_{as}\{n^{-1/2}(\hat{\lambda}_j - \lambda_j)\} = \frac{2\lambda_j R}{P_j} \qquad \text{for } j = 0,1,\ldots,M-1 \qquad (13)$$

$$V_{as}\{n^{-1/2}(\hat{\mu}_j - \mu_j)\} = \frac{2\mu_j R}{P_j} \qquad \text{for } j = 1,2,\ldots,M \qquad (14)$$

Suppose we hypothesize that $\underline{\theta}$ is some particular vector, $\underline{\theta}^o = (\lambda_0^o,\ldots,\lambda_{M-1}^o,\mu_1^o,\ldots,\mu_M^o)'$. If $H_0 : \underline{\theta} = \underline{\theta}^o$ is true, then by Theorem 3,

$$2[\max_{\underline{\theta}} \ln L_n(\underline{\theta}) - \ln L_n(\underline{\theta}^o)] \xrightarrow{L} \chi^2_{2M} \quad ,$$

where $2M$ is the number of parameters (components) of $\underline{\theta}$. For this model, we obtain:

$$2[\max_{\underline{\theta}} \ln L_n(\underline{\theta}) - \ln L_n(\underline{\theta}^o)] = \sum_{j=0}^{M-1} u_j \ln(u_j/\lambda_j^o r_j) + \sum_{j=1}^{M} d_j \ln(d_j/\mu_j^o r_j)$$

$$+ \sum_{j=0}^{M} r_j(\lambda_j^o + \mu_j^o) - n$$

which is computable from the sample. We determine a critical region such that

$P(\chi^2_{2M} \geq c) = \alpha$, then reject $H_0 : \underline{\theta} = \underline{\theta}^o$ if the number obtained from (15) exceeds c .

If H_0 is false such that $\underline{\theta} = \underline{\theta}^1$ say, then the power of the test can be determined as a function of $\underline{\theta}^o$ and $\underline{\theta}^1$. For $\underline{h} = n^{1/2}(\underline{\theta}^o - \underline{\theta}^1)$, the quantity $\Delta^2 = \underline{h}'\sigma(\underline{\theta}^1)\underline{h}$ is easily computed:

$$\Delta^2 = \frac{n}{2R}\left[\frac{P_0(\lambda_0^o - \lambda_0^1)^2}{\lambda_0^1} + \ldots + \frac{P_{M-1}(\lambda_{M-1}^o - \lambda_{M-1}^1)^2}{\lambda_{M-1}^1} + \frac{P_1(\mu_1^o - \mu_1^1)^2}{\mu_1^1} + \ldots + \frac{P_M(\mu_M^o - \mu_M^1)^2}{\mu_M^1}\right] \qquad (16)$$

From Theorem 4 we have:

$$2[\operatorname*{Max}_{\underline{\theta}} \ln L_n(\underline{\theta}) - \ln L_n(\underline{\theta}^o)] \xrightarrow{L} \chi^2_{2M,\Delta}$$

Given $\underline{\theta}^o$ and $\underline{\theta}^1$, Δ can be computed from (16), and then the power of the test determined as $P(\chi^2_{2M,\Delta} > c)$.

Nested within this model is the simple queue limit model where it is assumed that λ_j and μ_j do not vary with the number in the system. Suppose one owns a gas station and has an opportunity to enlarge his facility by buying adjacent land. How much additional business will he get? Does he lose business only when the facility is full (state E_M) , or are customers also discouraged by queue sizes less than $M - 1$? Without assuming any relationship among the μ_j , the hypothesis that $\lambda_0 = \lambda_1 = \ldots = \lambda_{M-1} = \lambda$ can be tested, where M is the maximum number of cars the facility can hold. Let $\underline{\phi}' = (\lambda, \mu_1, \mu_2, \ldots, \mu_M)$.

From (10), the likelihood of the restricted model is:

$$\ln L_n(\underline{\phi}) = \ln\lambda \sum_{j=0}^{M-1} u_j + \sum_{j=1}^{M} d_j \ln\mu_j - \sum_{j=0}^{M} \tau_j[(1 - \delta_{jM})\lambda + \mu_j]$$

and

$$\frac{\partial}{\partial\lambda} \ln L_n(\underset{\sim}{\phi}) = \sum_{j=0}^{M-1} u_j/\lambda - (T - \gamma_M) = 0$$

therefore

$$\hat{\lambda} = \frac{\sum_{j=0}^{M-1} u_j}{T - \gamma_M}$$

$$\frac{\partial}{\partial\mu_j} \ln L_n(\underset{\sim}{\phi}) = d_j/\mu_j - \gamma_j = 0 \qquad j = 1,2,\ldots,M$$

therefore

$$\hat{\mu}_j = \frac{d_j}{\gamma_j} \quad .$$

If the restricted hypothesis: $\lambda_0 = \ldots = \lambda_{M-1} = \lambda$ is true, then by Theorem 5,

$$2[\underset{\underset{\sim}{\theta}}{\text{Max}} \ln L_n(\underset{\sim}{\theta}) - \underset{\underset{\sim}{\phi}}{\text{Max}} \ln L_n(\underset{\sim}{\phi})] \xrightarrow{L} \chi^2_{k-k'} = \chi^2_{M-1}$$

where

$$2[\underset{\underset{\sim}{\theta}}{\text{Max}} \ln L_n(\underset{\sim}{\theta}) - \underset{\underset{\sim}{\phi}}{\text{Max}} \ln L_n(\underset{\sim}{\phi})] = 2\left[\sum_{j=0}^{M-1} u_j \ln\left(\frac{u_j}{\gamma_j\hat{\lambda}}\right) + \hat{\lambda}(T - \gamma_M) - \sum_{j=0}^{M-1} u_j\right] \quad (17)$$

which is easily computable from the sample.

MODEL 2: $\lambda_j = ah(j) \qquad j = 0,1,\ldots \qquad$ and $\qquad g(0) = 0$

$\mu_j = bg(j)$

where a and b are unknown parameters, and $h(j)$ and $g(j)$ are assumed to be arbitrary but known functions of j .

The advantage of the model is that it is not necessary to impose a queue limit in order to keep the number of parameters finite. Furthermore, information about a and b is obtained throughout T , rather than only when the system is in certain states. Hence, statistical inferences will be more precise. However, in order to use such a model, considerable information about the behavior of the system must already be at hand. When a queue limit is imposed by the nature of the system, then Model 1 is more general than Model 2.

For Model 2, (3) takes on the following form:

$$\ln L_n(\underline{\theta}) = \sum_{j=0}^{\infty} u_j \ln ah(j) + \sum_{j=1}^{\infty} d_j \ln bg(j) - \sum_{j=0}^{\infty} \tau_j[ah(j) + bg(j)] \qquad (18)$$

from which we obtain:

$$\hat{a} = \sum_{j=0}^{\infty} u_j \Big/ \sum_{j=0}^{\infty} \tau_j h(j) \qquad (19)$$

$$\hat{b} = \sum_{j=1}^{\infty} d_j \Big/ \sum_{j=1}^{\infty} \tau_j g(j) \quad . \qquad (20)$$

Note that $\lim_{T\to\infty} \sum_{j=0}^{\infty} u_j \Big/ \sum_{j=1}^{\infty} d_j = 1$ with probability one, for in ergodic queues, whatever goes up must come down! For large samples, we can therefore replace $\sum_{j=0}^{\infty} u_j$ and $\sum_{j=1}^{\infty} d_j$ by $n/2$.

We now obtain the variance-covariance matrix, starting again with the likelihood of a single transition:

$$dF(x_{i+1},\tau_i \mid x_i = j,\underline{\theta}) = ah(j)\ \exp -(ah(j)+bg(j))\tau_i \qquad \text{if} \quad x_{i+1} = j+1$$

$$= bg(j)\ \exp -(ah(j)+bg(j))\tau_i \qquad \text{if} \quad x_{i+1} = j-1$$

and given $x_i = j$,

$$G_a = \frac{1}{a} - h(j)\tau_i \quad , \quad G_b = \quad - g(j)\tau_i \qquad \text{if} \quad x_{i+1} = j+1$$

$$G_a = \quad - h(j)\tau_i \quad , \quad G_b = \frac{1}{b} - g(j)\tau_i \qquad \text{if} \quad x_{i+1} = j-1$$

$$G_{aa} = -\frac{1}{a^2} \qquad G_{bb} = 0 \qquad \text{if} \quad x_{i+1} = j+1$$

$$G_{aa} = 0 \qquad G_{bb} = -\frac{1}{b^2} \qquad \text{if} \quad x_{i+1} = j-1$$

$$G_{ab} = 0 \quad .$$

Hence:

$$V(G_a \mid x_i = j) = (1/a^2)\ \frac{ah(j)}{ah(j) + bg(j)} = \frac{h(j)}{a[ah(j) + bg(j)]}$$

$$V(G_b \mid x_i = j) = (1/b^2)\ \frac{bg(j)}{ah(j) + bg(j)} = \frac{g(j)}{b[ah(j) + bg(j)]}$$

Cov $(G_a,G_b \mid x_i = j) = 0$ and

$$V(G_a) = \sum_{j=0}^{\infty} V(G_a \mid x_i = j)\pi_j = \sum_{j=0}^{\infty} \frac{h(j)\pi_j}{a[ah(j) + bg(j)]}$$

$$V(G_b) = \sum_{j=0}^{\infty} V(G_b \mid x_i = j)\pi_j = \sum_{j=0}^{\infty} \frac{g(j)\pi_j}{b[ah(j) + bg(j)]} \quad .$$

Now from (4), $\pi_j = P_j[ah(j) + bg(j)]/2R$, and we obtain the simple result:

$$V(G_a) = \frac{1}{2Ra} \sum_{j=0}^{\infty} h(j)P_j = \frac{1}{2a^2}$$

$$V(G_b) = \frac{1}{2Rb} \sum_{j=0}^{\infty} g(j)P_j = \frac{1}{2b^2}$$

$$\text{Cov}\ (G_a, G_b) = 0 \quad .$$

For $\underline{\theta}' = (a,b)$,

$$\sigma(\underline{\theta}) = \begin{pmatrix} 1/2a^2 & 0 \\ 0 & 1/2b^2 \end{pmatrix}$$

again a diagonal matrix.

Thus for $n^{1/2}(\hat{\underline{\theta}} - \underline{\theta})' = n^{1/2}(\hat{a} - a\ ,\ \hat{b} - b)$, $n^{1/2}(\hat{\underline{\theta}} - \underline{\theta}) \xrightarrow{L} N(\underline{0}\ ,\ \sigma^{-1}(\underline{\theta}))$ where

$$\sigma^{-1}(\underline{\theta}) = \begin{pmatrix} 2a^2 & 0 \\ 0 & 2b^2 \end{pmatrix} \tag{21}$$

Note that the form of this result is independent of $h(j)$ and $g(j)$. However, this does not imply that the variances are independent of $h(j)$ and $g(j)$, since for given T , the distribution of n depends on them.

Hypotheses about a and b can also be tested by the same methods used in Model 1. If $\underline{\theta}^0 = (a^0, b^0)'$ is hypothesized, but $\underline{\theta}^1 = (a^1, b^1)'$ is true, such that $\underline{h} = \sqrt{n}(\underline{\theta}^0 - \underline{\theta}^1)$, then:

$$\Delta^2 = \underline{h}'\sigma(\underline{\theta})\ \underline{h} = \frac{n}{2}\left[\left(\frac{a^o - a^1}{a^1}\right)^2 + \left(\frac{b^o - b^1}{b^1}\right)^2\right] \tag{22}$$

and

$$2\,[\underset{\underline{\theta}}{\text{Max}}\ \ln L_n\ (\underline{\theta}) - \ln L_n(\underline{\theta}^o)] \rightarrow \chi^2_{2,\Delta} \tag{23}$$

<u>Special Cases of Model 2 Which are of Independent Interest</u>: Included in Model 2 are many common queueing models. Some of these are listed below:

<u>Model 2a</u>: The infinite channel model, $\lambda_j = \lambda$, $\mu_j = j\mu$, for $j = 0,1,\ldots$. This is simply Model 2 with $\lambda = a, h(j) = 1$, $b = \mu$, $g(j) = j$.

<u>Model 2b</u>: The C-channel model, $\lambda_j = \lambda$, $\mu_j = j\mu$ for $j = 0,1,\ldots,C-1$ $\mu_j = C\mu$ for $j = C,C+1,\ldots$. This also is Model 2 with $\lambda = a, h(j) = 1$, $b = \mu$, and $g(j) = j$ for $j \leq C - 1$, $g(j) = C$ for $j \geq C$.

<u>Model 2c</u>: The machine-repair model, single repairman case. Suppose there are M machines, each of which break down at rate λ when operating. The machines are in operation continuously except when they break down. The repairman works on one machine at a time, at rate μ . Let the state of the system, E_j , be defined as the number of working machines. Then we have: $\lambda_j = j\lambda$ for $j \leq M$, $\mu_j = \mu$ for $j \leq M-1$, $\mu_M = 0$. Note that a service corresponds to a transition $E_j \rightarrow E_{j+1}$ while an arrival results in a transition $E_j \rightarrow E_{j-1}$. Thus we have a truncated C-channel model with the roles of λ_j and μ_j interchanged. Note also that this model is nested in Model 1 .

<u>Model 2d</u>: Model 2c with N reparimen such that $N < M$, and each repairman works alone at rate μ . For this case: $\lambda_j = j\lambda$ for $0 \leq j \leq M$, $\mu_j = (M-1)\mu$ for $M - N \leq j \leq M$, $\mu_j = N\mu$ for $0 \leq j \leq M - N$. This model is also nested in Model 1, and represents the hypothesis that the rate repairmen work is not affected by the number of jobs waiting. Assuming Model 1 correctly represents the system, this hypothesis can be tested by the methods presented earlier.

Rejection of the hypothesis leads to the conclusion that the number of jobs waiting does affect the performance of repairmen. It would be more efficient, however, to nest this model in a model more restrictive than Model 1, assuming that the new model is true. One such model is: $\lambda_j = j\lambda$, and $\mu_j = \mu_j$ for $0 \leq j \leq M$.

Comparing Model 1 and Model 2, we see that Model 1 is more general, but if the queue limit must be made large to be realistic, the number of parameters becomes correspondingly large. Under these conditions, tests lose power, and particularly for large j , the variances of estimators are apt to be large. On the other hand, Model 2 is rather restrictive and requires a lot of a priori information to be useful. A compromise is suggested. Model 3 is such a compromise.

MODEL 3: $\lambda_j > 0$ for $j = 0,\ldots,M-1$, $\mu_j > 0$ for $j = 1,\ldots,M-1$

$\lambda_j = ah(j)$, $\mu_j = bg(j)$ for $j \geq M$

For situations where large queues are possible, but the service facility performs well, relative to the load, such that the average queue length is small, this model will be quite useful. The λ_j and μ_j can be individually estimated for small j , where the system is most of the time, and lumped together for large j , without serious consequences.

From (3), we obtain for this model:

$$\hat{\lambda}_j = u_j/\gamma_j \ , \ \hat{\mu}_j = d_j/\gamma_j \tag{25}$$

as before, and

$$\hat{a} = \sum_{j=M}^{\infty} u_j \Big/ \sum_{j=M}^{\infty} \gamma_j h(j) \tag{26}$$

$$\hat{b} = \sum_{j=M}^{\infty} d_j \Big/ \sum_{j=M}^{\infty} r_j g(j) \tag{27}$$

In obtaining the variance-covariance matrix, all the covariances will again be zero. The only part that is at all new is the determination of $V(G_a)$ and $V(G_b)$. If $x_i = j < M$, G_a and G_b are identically zero. Hence we need only consider $x_i = j \geq M$.

Given $x_i = j \geq M$

$$G_a = \frac{1}{a} - h(j)\tau_i \quad , \quad G_b = -g(j)\tau_i \qquad \text{if } x_{i+1} = j + 1$$

$$G_a = -h(j)\tau_i \qquad G_b = \frac{1}{b} - g(j)\tau_i \qquad \text{if } x_{i+1} = j - 1 \ .$$

It is now clear that:

$$V(G_a) = \sum_{j=M}^{\infty} \frac{h(j)\pi_j}{a[ah(j) + bg(j)]} = \frac{1}{2Ra} \sum_{j=M}^{\infty} h(j)P_j$$

$$V(G_b) = \sum_{j=M}^{\infty} \frac{g(j)\pi_j}{b[ah(j) + bg(j)]} = \frac{1}{2Rb} \sum_{j=M}^{\infty} g(j)P_j$$

and

$$V_{as}\{\sqrt{n}(\hat{a} - a)\} = 2Ra \Big/ \sum_{j=M}^{\infty} h(j)P_j \tag{28}$$

$$V_{as}\{\sqrt{n}(\hat{b} - b)\} = 2Rb \Big/ \sum_{j=M}^{\infty} g(j)P_j \tag{29}$$

Many intereting possibilities are uncovered by noting that Model 2 is nested in Model 3. Thus, one can test the hypothesis that a particular $h(j)$ and or $g(j)$ applied to the lower states only. The number of states included in such a test is determined by M. Further, there is no need to require that M be the same for both the $\{\lambda_j\}$ and the $\{\mu_j\}$.

VIII. An Example Using Data Generated by Simulation

We would like to know how well the theory, when applied to models such as those in section VII, can discriminate between those which reflect the structure of a system and those which do not. An analysis of data generated by simulating a model of known structure will be of value here.

For the simulation, the following model was used:

$$\lambda_j = 4/(j+1) \qquad j = 0,1,\ldots$$
$$\mu_0 = 0$$
$$\mu_1 = 1$$
$$\mu_j = 2 \qquad j = 2,3,\ldots$$

Results of the simulation (along with P_j(true) for comparison) were as follows:

j	μ_j	d_j	γ_j	$\hat{P}_j = \gamma_j/\Sigma\gamma_j$	P_j(true)
0	44	0	10.54	.065	.073
1	93	44	50.76	.312	.291
2	63	93	39.33	.242	.291
3	31	63	32.34	.198	.193
4	13	31	15.90	.098	.097
5	5	13	10.66	.066	.039
6	1	5	2.54	.016	.013
7	0	1	.73	.004	.004
Σ	250	250	162.80	1.001	1.001

Note that $u_j = d_{j+1}$. This relationship must hold for all states except initial and final, and, by chance, did for them too in this simulation.

It is not difficult to show that $\sum_{j=8}^{\infty} P_j < .02\, P_0$ and hence is negligible. We also compute the true throughput, $R = \mu_1 P_1 + \mu_2(1 - P_0 - P_1) =$ $= 1.56$.

Assume for concreteness that we have a 2-channel facility for which it has been observed that rarely are there more than 4 customers in the system at any time. (It might be a gas station, a bank, or even a bordello.) We would like to know how λ_j varies with j. Specifically, for small j, does it decline with j or remain constant? The economics of enlarging the facility are closely tied to this relationship.

The model we will assume in the analysis is the following:

Model M_1 : H_0 : $\lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4 > 0$

$$\lambda_j = a/(j + 1) \qquad j = 5,6,\ldots$$

$$\mu_0 = 0$$

$$\mu_1 = \mu$$

$$\mu_j = 2\mu \qquad j = 2,3,\ldots$$

Note that there are 7 unknown parameters.

Estimation of parameters and their coefficients of variation

	true θ	est. eq.	$\hat{\theta}$	V_{as} eq.	$V_{as}(\hat{\theta} - \theta)$	σ/μ
λ_0	4.00	(25)	4.17	(13)	.342	.146
λ_1	2.00	(25)	1.83	(13)	.0428	.103
λ_2	1.33	(25)	1.60	(13)	.0285	.126
λ_3	1.00	(25)	0.96	(13)	.0321	.179
λ_4	0.80	(25)	0.82	(13)	.0499	.280
a	4.00	(26)	2.68	(28)	2.81	.418
μ	1.00	(20)	0.99	(21)	.0040	.064

Let us denote the estimate of a in this table by $\hat{a}_1 = 2.68$.

Next consider another model,

Model M_2 : H_0 : $\lambda_j = a/(j + 1)$ $\qquad j = 0,1,\ldots$

$$\mu_0 = 0$$

$$\mu_1 = \mu$$

$$\mu_j = 2\mu \qquad j = 2,3,\ldots$$

(this assumption is correct) and we obtain:

$$\hat{a}_2 = 4.00$$

$$V_{as}(\hat{a}_2 - a) = .064$$

$$(\sigma/\mu)_{\hat{a}_2} = .064 = (\sigma/\mu)_{\hat{\lambda}_j} \quad \text{as well.}$$

We will now test whether model M_2 is correct noting that it is nested in M_1 . There were 7 unknown parameters under the original assumption, while there are two, $\phi' = (a,\mu)$ under hypothesis M_2 . Hence, from Theorem 5 we obtain:

$$2[\operatorname*{Max}_{\underline{\theta}} \ln L_n(\underline{\theta}) - \operatorname*{Max}_{\phi} \ln L_n(\phi)] \xrightarrow{L} \chi^2_{7-2} \quad \text{if } H_0 \text{ is true.}$$

From (3) and the maximum likelihood estimators, the L.H.S. becomes:

$$2\left[\sum_{j=0}^{4} u_j \ln\hat{\lambda}_j + \sum_{j=5}^{7} u_j \ln(\hat{a}_1/(j + 1)) - \sum_{j=0}^{4} u_j - \sum_{j=5}^{7} \hat{a}_1 r_j/(j + 1)\right.$$

$$\left. - \sum_{j=0}^{7} u_j \ln(\hat{a}_2/(j + 1)) + \sum_{j=0}^{7} \hat{a}_2 r_j/(j + 1)\right]$$

$$= 3.32 < \chi^2_{5,.05} = 11.07 \quad .$$

Hence, we would accept this hypothesis at any reasonable α level.

On the other hand, let us test an alternative hypothesis: Model M_3 : H_0 : $\lambda_0 = \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda$, unrelated to a , with the same assumptions about the $\{\mu_j\}$. (This model is not correct.) This hypothesis is also nested in the original one, but with three unknown parameters, $\underline{\phi}' = (\lambda, a, \mu)$. We have

$$2[\underset{\underline{\theta}}{\mathrm{Max}}\ \ln L_n(\underline{\theta}) - \underset{\underline{\phi}}{\mathrm{Max}}\ \ln L_n(\underline{\phi})] \xrightarrow{L} \chi^2_4 \quad .$$

Using $\hat{\lambda} = \sum_{j=0}^{4} u_j / \sum_{j=0}^{4} r_j = 1.640$, the L.H.S. becomes:

$$2\left[\sum_{j=0}^{4} u_j \ln(\hat{\lambda}_j/\hat{\lambda}) - \sum_{j=0}^{4} u_j + \sum_{j=0}^{4} r_j\hat{\lambda}\right] = 52 > \chi^2_{4,.05} = 9.49$$

$$> \chi^2_{4,.0005} = 20 \quad .$$

Hence, we would reject M_3 at any reasonable α level.

Although it is obvious from the raw data that M_3 is untenable, it should be clear that these methods can be applied to more complicated cases where inferences by inspection are hard to substantiate.

APPENDIX

IDEAS FOR FURTHER RESEARCH

Models of Greater Generality

These ideas can be extended by partitioning the $\{E_j\}$ into subsets, $U_1, U_2, \ldots, U_m$, such that if $E_j \in U_i$, $\lambda_j = a_i h_i(j)$. Similarly, partition $\{E_j\}$ into $D_1, D_2, \ldots, D_n$ such that if $E_j \in D_i$, $\mu_j = b_i g_i(j)$. If all the parameters are independent, one obtains:

$$\hat{a}_i = \sum_{E_j \in U_i} u_j \Big/ \sum_{E_j \in U_i} \gamma_j h_j(j) \tag{A-1}$$

$$\hat{b}_i = \sum_{E_j \in D_i} d_j \Big/ \sum_{E_j \in D_i} \gamma_j g_i(j) \tag{A-2}$$

$$V_{as}\{\sqrt{n}(\hat{a}_i - a_i)\} = 2Ra_i \Big/ \sum_{E_j \in U_i} h_i(j) P_j \tag{A-3}$$

$$V_{as}\{\sqrt{n}(\hat{b}_i - b_i)\} = 2Rb_i \Big/ \sum_{E_j \in D_i} g_i(j) P_j \tag{A-4}$$

All covariances are zero.

In all the models considered thus far, all the covariances have been zero, so that $\sigma(\underline{\theta})$ has been a diagonal matrix. Now we will investigate why this has been so.

Sufficient Conditions for $\sigma(\theta)$ to be a Diagonal Matrix

The matrix $\sigma(\underline{\theta})$ is diagonal $\Longleftrightarrow$ $E(\sigma_{uv}) = 0$ for every (u,v) such that

$u \neq v \Leftrightarrow$ the covariances of the $\{G_u\}$ are all zero. The result is desirable in that the variance-covariance matrix of the estimators, $\sigma^{-1}(\underline{\theta})$, is simple to find and is also diagonal. The estimators, being already asymptotically normal, are then also asymptotically independent. Results can be stated in a simple form.

Now $E(G_{uv}) = E\{E(G_{uv} \mid x_i = j)\}$. Hence, $E(G_{uv}) = 0$ if $E(G_{uv} \mid x_i = j) = 0$ for all j . This will occur if $dF_i \mid x_i$ is separable in the parameters, or equivalently if $\ln dF_i \mid x_i$ is additive in the parameters. We say $dF_i \mid x_i$ is separable in $\{\theta_u\}$ if $dF_i \mid x_i$ can be expressed in the following form:

$$dF_i \mid x_i = \prod_u g_u(x_{i+1} , \tau_i \mid x_i , \theta_u)$$

so that $g_v(x_{i+1} , \tau_i \mid x_i , \theta_v)$ does not involve u unless $u = v$. If this is the case, then:

$$\ln dF_i \mid x_i = \sum_u \ln g_u(x_{i+1} , \tau_i \mid x_i , \theta_u)$$

and we say $\ln dF_i \mid x_i$ is additive in the parameters. When $dF_i \mid x_i$ is separable in the θ_u , then for

$$u \neq v , \; G_{uv} \mid x_i = \frac{\partial^2 \ln dF_i}{\partial\theta_u \partial\theta_v} \equiv 0 \Rightarrow E(G_{uv} \mid x_i) = 0 \Rightarrow E(G_{uv}) = 0 \quad .$$

For all the models consideded thus far, arrival and service parameters occur together only in the exponent, and there as a sum. Hence, they are separable. Further, given the state of the system, at most one arrival and one service parameter appears in dF_i .

Not only have the likelihood functions been separable, but the corresponding models have been linear. That is, the arrival and service rates have been linear functions of the unknown parameters. The analysis of variance and regression are based on linear models, which are surprisingly adaptable in their ability to cover many situations. Further, deviating from linearity here results in considerable algebraic complexity. Thus, only linear models have been considered. Deviating from separability also increases complexity, but to a lesser degree. The nonseparable linear model is considered briefly below.

<u>General Linear Model</u>

$$\lambda_j = \sum_{y=1}^{Y} a_y h_y(j)$$

$$\mu_j = \sum_{z=1}^{Z} b_z g_z(j)$$

where $g_z(0) = 0$ for all z. From (2.3) we obtain:

$$\ln L_n(\underline{\theta}) = \sum_{j=0}^{\infty} u_j \ln\left[\sum_{y=1}^{Y} a_y h_y(j)\right] + \sum_{j=1}^{\infty} d_j \ln\left[\sum_{z=1}^{Z} b_z g_z(j)\right] - \sum_{j=0}^{\infty} \gamma_j \left[\sum_{y=1}^{Y} a_y h_y(j) + \sum_{z=1}^{Z} b_z g_z(j)\right] \qquad \text{(A-5)}$$

and

$$\frac{\partial \ln L_n(\underline{\theta})}{\partial a_y} = \sum_{j=0}^{\infty} \frac{u_j h_y(j)}{\sum_{y=1}^{Y} a_y h_y(j)} - \sum_{j=0}^{\infty} \gamma_j h_y(j) = 0 \quad \text{for} \quad y = 1,2,\ldots,Y$$

$$\frac{\partial \ln L_n(\underline{\theta})}{\partial b_z} = \sum_{j=0}^{\infty} \frac{d_j g_z(j)}{\sum_{z=1}^{Z} b_z g_z(j)} - \sum_{j=0}^{\infty} \tau_j g_z(j) = 0 \quad \text{for} \quad z = 1,2,\ldots,Z \quad .$$

Although the above equations will have to be solved numerically to obtain the maximum likelihood estamators, the variance-covariance matrix of the asymptotic distribution is relatively easy to find. Below we find $\sigma(\underline{\theta})$

Given $x_i = j$,

$$G_{a_y} = \frac{h_y(j)}{\sum_{y=1}^{Y} a_y h_y(j)} - h_y(j)\tau_i \qquad \text{if} \quad x_{i+1} = j + 1$$

$$= \qquad - h_y(j)\tau_i \qquad \text{if} \quad x_{i+1} = j - 1$$

$$G_{b_z} = -b_z(j)\tau_i \qquad \text{if} \quad x_{i+1} = j + 1$$

$$= \frac{b_z(j)}{\sum_{z=1}^{Z} b_z g_z(j)} - b_z(j)\tau_i \qquad \text{if} \quad x_{i+1} = j - 1$$

therefore

$$V(G_{a_y} \mid x_i = j) = \frac{h_y^2(j)}{\left(\sum_{y=1}^{Y} a_y h_y(j)\right)\left(\sum_{y=1}^{Y} a_y h_y(j) + \sum_{z=1}^{Z} b_z g_z(j)\right)}$$

and from:

$$\pi_j = \frac{P_j\left(\sum_{y=1}^{Y} a_y h_y(j) + \sum_{z=1}^{Z} b_z g_z(j)\right)}{2R}$$

we obtain

$$V(G_{a_y}) = \frac{1}{2R} \sum_{j=0}^{\infty} \frac{h_y^2(j)P_j}{\sum_{y=1}^{Y} a_y h_y(j)} .$$

Similarly,

$$\mathrm{Cov}(G_{a_y}, G_{a_z} \mid x_i = j) = \frac{h_y(j)h_z(j)}{\sum_{y=1}^{Y} a_y h_y(j) \quad \sum_{y=1}^{Y} a_y h_y(j) + \sum_{z=1}^{Z} b_z g_z(j)}$$

therefore

$$\mathrm{Cov}(G_{a_y}, G_{a_z}) = \frac{1}{2R} \sum_{j=0}^{\infty} \frac{h_y(j)h_z(j)P_j}{\sum_{y=1}^{Y} a_y h_y(j)}$$

Similar expressions may be obtained for the G_{b_k} . Note that $\mathrm{Cov}(G_{a_y}, G_{b_z}) = 0$.

Queueing Processes of Greater Generality

By restricting the class of queueing processes in this paper, it has been possible to apply work done in Markov processes to problems of interest in queueing. The author has also obtained results for certain more general queueing processes, those which possess embedded Markov chains [7]. One important process of this type is single channel with Poisson arrivals and arbitrary service. This work will be published at a later date.

REFERENCES

1. Billingsley, P., Statistical Inference for Markov Processes, University of Chicago Press, Chicago, 1961.

2. Billingsley, P., "Statistical Methods in Markov Chains," Ann. Math. Stat., Vol. 32, No. 1, (1961).

3. Clark, A. B., "Maximum Likelihood Estimates in a Simple Queue," Ann. Math. Stat., Vol. 28, No. 4 (1957).

4. Cramer, H. C., Mathematical Methods of Statistics, Princeton University Press, Princeton, 1946.

5. Feller, W., An Introduction of Probability Theory and its Applications, 2nd Ed., John Wiley and Sons, New York, 1957.

6. Wilks, S. S., Mathematical Statistics, John Wiley and Sons, New York, 1962.

7. Wolff, R. W., Statistical Inference In Queueing Processes, Ph. D. Thesis, Case Institute of Technology, 1963. (Available through University Microfilms, Inc, Ann Arbor, Michigan).